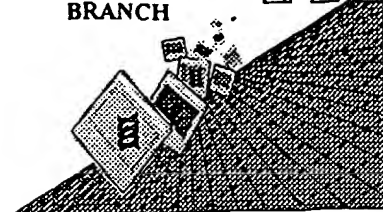




RAW SEQUENCE LISTING ERROR REPORT



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/830,026

Source: PCT 09

Date Processed by STIC: 5-7-01

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 - 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO).

Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

New Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/830,026

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics The number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".

- 2 Wrapped Aminos The amino acid number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".

- 3 Incorrect Line Length The rules require that a line not exceed 72 characters in length. This includes spaces.

- 4 Misaligned Amino Acid The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs
Numbering between the numbering. It is recommended to delete any tabs and use spacing between the numbers.

- 5 Non-ASCII This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.

- 6 Variable Length Sequence(s) contain n's or Xaa's which represented more than one residue.
As per the rules, each n or Xaa can only represent a single residue.
Please present the maximum number of each residue having variable length and
indicate in the (ix) feature section that some may be missing.

- 7 PatentIn ver. 2.0 "bug" A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid
sequence(s) . Normally, PatentIn would automatically generate this section from the
previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section
to the subsequent amino acid sequence. This applies primarily to the mandatory <220>-<223>
sections for Artificial or Unknown sequences.

- 8 Skipped Sequences Sequence(s) missing. If intentional, please use the following format for each skipped sequence:
(OLD RULES) (2) INFORMATION FOR SEQ ID NO:X:
(i) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:X:
This sequence is intentionally skipped

Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).

- 9 Skipped Sequences Sequence(s) missing. If intentional, please use the following format for each skipped sequence.
(NEW RULES) <210> sequence id number
 <400> sequence id number
 000

- 10 Use of n's or Xaa's Use of n's and/or Xaa's have been detected in the Sequence Listing.
(NEW RULES) Use of <220> to <223> is MANDATORY if n's or Xaa's are present.
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.

- 11 ✓ Use of "Artificial" Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules.
(NEW RULES) Valid response is Artificial Sequence.

- 12 Use of <220>Feature Sequence(s) are missing the <220>Feature and associated headings.
(NEW RULES) Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial Sequence" or "Unknown"
Please explain source of genetic material in <220> to <223> section.
(See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of new Rules)

- 13 PatentIn ver. 2.0 "bug" Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted
file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing).
Instead, please use "File Manager" or any other means to copy file to floppy disk.

PCT

RAW SEQUENCE LISTING DATE: 05/07/2001
 PATENT APPLICATION: US/09/830,026 TIME: 14:25:59

Input Set : A:\Uok532-1.txt
 Output Set: N:\CRF3\05072001\I830026.raw

3 <110> APPLICANT: University of Kansas Center for Research
 4 Walter Reed Army Institute for Research
 6 <120> TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF PURIFIED INVASIN PROTEIN AND USE THEREOF
 8 <130> FILE REFERENCE: UOK 5320.1
 C--> 10 <140> CURRENT APPLICATION NUMBER: US/09/830,026
 C--> 10 <141> CURRENT FILING DATE: 2001-04-20
 10 <150> PRIOR APPLICATION NUMBER: PCT/US99/24931
 11 <151> PRIOR FILING DATE: 1999-10-21
 13 <160> NUMBER OF SEQ ID NOS: 17
 15 <170> SOFTWARE: PatentIn version 3.0
 17 <210> SEQ ID NO: 1
 18 <211> LENGTH: 409
 19 <212> TYPE: PRT
 20 <213> ORGANISM: Salmonella typhimurium
 22 <400> SEQUENCE: 1
 24 Met Leu Ile Ser Asn Val Gly Ile Asn Pro Ala Ala Tyr Leu Asn Asn
 25 1 5 10 15
 27 His Ser Val Glu Asn Ser Ser Gln Thr Ala Ser Gln Ser Val Ser Ala
 28 20 25 30
 30 Lys Asp Ile Leu Asn Ser Ile Gly Ile Ser Ser Ser Lys Val Ser Asp
 31 35 40 45
 33 Leu Gly Leu Ser Pro Thr Leu Ser Ala Pro Ala Pro Gly Val Leu Thr
 34 50 55 60
 36 Gln Thr Pro Gly Thr Ile Thr Ser Ser Leu Lys Ala Ser Ile Gln Asn
 37 65 70 75 80
 39 Thr Asp Met Asn Gln Asp Leu Asn Ala Leu Ala Asn Asn Val Thr Thr
 40 85 90 95
 42 Lys Ala Asn Glu Val Val Gln Thr Gln Leu Arg Glu Gln Gln Ala Glu
 43 100 105 110
 45 Val Gly Lys Phe Phe Asp Ile Ser Gly Met Ser Ser Ser Ala Val Ala
 46 115 120 125
 48 Leu Leu Ala Ala Ala Asn Thr Leu Met Leu Thr Leu Asn Gln Ala Asp
 49 130 135 140
 51 Ser Lys Leu Ser Gly Lys Leu Ser Leu Val Ser Phe Asp Ala Ala Lys
 52 145 150 155 160
 54 Thr Thr Ala Ser Ser Met Met Arg Glu Gly Met Asn Ala Leu Ser Gly
 55 165 170 175
 57 Ser Ile Ser Gln Ser Ala Leu Gln Leu Gly Ile Thr Gly Val Gly Ala
 58 180 185 190
 60 Lys Leu Glu Tyr Lys Gly Leu Gln Asn Glu Arg Gly Ala Leu Lys His
 61 195 200 205
 63 Asn Ala Ala Lys Ile Asp Lys Leu Thr Thr Glu Ser His Ser Ile Lys
 64 210 215 220
 66 Asn Val Leu Asn Gly Gln Asn Ser Val Lys Leu Gly Ala Glu Gly Val
 67 225 230 235 240
 69 Asp Ser Leu Lys Ser Leu Asn Ile Arg Lys Pro Val Pro Met Arg Arg
 70 245 250 255

Does Not Comply
 Corrected Diskette Needed
 pp. 3-5

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/830,026

DATE: 05/07/2001

TIME: 14:25:59

Input Set : A:\Uok532-1.txt

Output Set: N:\CRF3\05072001\I830026.raw

```

72 Lys Ile Leu Met Met Arg Arg Leu Asn Leu Met Pro Glu Pro Ala Pro
73      260      265      270
75 Arg Lys Val Trp Val Leu Lys Thr Val Ile Asn Lys Val Ser Leu Asn
76      275      280      285
78 Ile Tyr Ile Leu Ser Lys Arg Leu Glu Ser Val Glu Ser Asp Ile Arg
79      290      295      300
81 Leu Glu Gln Asn Tyr Met Asp Ile Thr Arg Ile Asp Ser Ala Gln Asp
82 305      310      315      320
84 Ala Asp Asp Gly Arg Ser Asp Tyr Glu Glu Leu Gly His Gly Arg Trp
85      325      330      335
87 Tyr Cys Arg Gly Val Arg Ala Val Arg Arg Tyr Ser Gly Asn Val Ser
88      340      345      350
90 Glu Gln Gln Ile Ser Gln Val Asn Asn Arg Val Ala Ser Thr Ala Ser
91      355      360      365
93 Asp Glu Ala Arg Glu Ser Ser Arg Lys Ser Thr Ser Leu Ile Gln Glu
94      370      375      380
96 Met Leu Lys Thr Met Glu Ser Ile Asn Gln Ser Lys Ala Ser Ala Leu
97 385      390      395      400
99 Ala Ala Ile Ala Gly Asn Ile Arg Ala
100      405
102 <210> SEQ ID NO: 2
103 <211> LENGTH: 382
104 <212> TYPE: PRT
105 <213> ORGANISM: Shigella flexneri
107 <400> SEQUENCE: 2
109 Met Leu Gln Lys Gln Phe Cys Asn Lys Leu Leu Leu Asp Thr Asn Lys
110 1      5      10      15
112 Glu Asn Val Met Glu Ile Gln Asn Thr Lys Pro Thr Gln Thr Leu Tyr
113      20      25      30
115 Thr Asp Ile Ser Thr Lys Gln Thr Gln Ser Ser Ser Glu Thr Gln Lys
116      35      40      45
118 Ser Gln Asn Tyr Gln Gln Ile Ala Ala His Ile Pro Leu Asn Val Gly
119      50      55      60
121 Lys Asn Pro Val Leu Thr Thr Thr Leu Asn Asp Asp Gln Leu Leu Lys
122 65      70      75      80
124 Leu Ser Glu Gln Val Gln His Asp Ser Glu Ile Ile Ala Arg Leu Thr
125      85      90      95
127 Asp Lys Lys Met Lys Asp Leu Ser Glu Met Ser His Thr Leu Thr Pro
128      100      105      110
130 Glu Asn Thr Leu Asp Ile Ser Ser Leu Ser Ser Asn Ala Val Ser Leu
131      115      120      125
133 Ile Ile Ser Val Ala Val Leu Ser Ala Leu Arg Thr Ala Glu Thr
134      130      135      140
136 Lys Leu Gly Ser Gln Leu Ser Leu Ile Ala Phe Asp Ala Thr Lys Ser
137 145      150      155      160
139 Ala Ala Glu Asn Ile Val Arg Gln Gly Leu Ala Ala Leu Ser Ser Ser
140      165      170      175
142 Ile Thr Gly Ala Val Thr Gln Val Gly Ile Thr Gly Ile Gly Ala Lys
143      180      185      190

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/830,026

DATE: 05/07/2001
TIME: 14:25:59

Input Set : A:\Uok532-1.txt
Output Set: N:\CRF3\05072001\I830026.raw

```

145 Lys Thr His Ser Gly Ile Ser Asp Gln Lys Gly Ala Leu Arg Lys Asn
146      195      200      205
148 Leu Ala Thr Ala Gln Ser Leu Glu Lys Glu Leu Ala Gly Ser Lys Leu
149      210      215      220
151 Gly Leu Asn Lys Gln Ile Asp Thr Asn Ile Thr Ser Pro Gln Thr Asn
152 225      230      235      240
154 Ser Ser Thr Lys Phe Leu Gly Lys Asn Lys Leu Ala Pro Asp Asn Ile
155      245      250      255
157 Ser Leu Ser Thr Glu His Lys Thr Ser Leu Ser Ser Pro Asp Ile Ser
158      260      265      270
160 Leu Gln Asp Lys Ile Asp Thr Gln Arg Arg Thr Tyr Glu Leu Asn Thr
161      275      280      285
163 Leu Ser Ala Gln Gln Lys Gln Asn Ile Gly Arg Ala Thr Met Glu Thr
164      290      295      300
166 Ser Ala Val Ala Gly Asn Ile Ser Thr Ser Gly Gly Arg Tyr Ala Ser
167 305      310      315      320
169 Ala Leu Glu Glu Glu Glu Gln Leu Ile Ser Gln Ala Ser Ser Lys Gln
170      325      330      335
172 Ala Glu Glu Ala Ser Gln Val Ser Lys Glu Ala Ser Gln Ala Thr Asn
173      340      345      350
175 Gln Leu Ile Gln Lys Leu Leu Asn Ile Ile Asp Ser Ile Asn Gln Ser
176      355      360      365
178 Lys Asn Ser Ala Ala Ser Gln Ile Ala Gly Asn Ile Arg Ala
179      370      375      380

```

```

181 <210> SEQ ID NO: 3
182 <211> LENGTH: 4
183 <212> TYPE: DNA
C--> 184 <213> ORGANISM: Artificial
186 <220> FEATURE:
187 <221> NAME/KEY: misc_feature
188 <222> LOCATION: (1)..(4)
189 <223> OTHER INFORMATION: NdeI restriction site
192 <400> SEQUENCE: 3
193 gaga
196 <210> SEQ ID NO: 4
197 <211> LENGTH: 29
198 <212> TYPE: DNA
C--> 199 <213> ORGANISM: Artificial
201 <220> FEATURE:
202 <221> NAME/KEY: misc_feature
203 <222> LOCATION: (1)..(29)
204 <223> OTHER INFORMATION: PCR Primer
207 <400> SEQUENCE: 4
208 gagacatatg ttatcagagc aggttcagc
211 <210> SEQ ID NO: 5
212 <211> LENGTH: 30
213 <212> TYPE: DNA
C--> 214 <213> ORGANISM: Artificial
216 <220> FEATURE:

```

*Circled <213> responses
are incomplete as per
section 1.8236 of new
sequence rules. See #11
on the Error Summary Sheet.*

*Note: This error is indicated
throughout the sequence
listing. Please review
and correct.*

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/830,026

DATE: 05/07/2001
TIME: 14:25:59

Input Set : A:\Uok532-1.txt
Output Set: N:\CRF3\05072001\I830026.raw

```
217 <221> NAME/KEY: misc_feature
218 <222> LOCATION: (1)..(30)
219 <223> OTHER INFORMATION: PCR Primer
222 <400> SEQUENCE: 5
223 gagaggatcc ttaagctcga atgttaccag 30
226 <210> SEQ ID NO: 6
227 <211> LENGTH: 27
228 <212> TYPE: DNA
C--> 229 <213> ORGANISM: Artificial
231 <220> FEATURE:
232 <221> NAME/KEY: misc_feature
233 <222> LOCATION: (1)..(27)
234 <223> OTHER INFORMATION: PCR Primer
237 <400> SEQUENCE: 6
238 gagacatatg ttgcaaaagc aatttgc 27
241 <210> SEQ ID NO: 7
242 <211> LENGTH: 32
243 <212> TYPE: DNA
C--> 244 <213> ORGANISM: Artificial
246 <220> FEATURE:
247 <221> NAME/KEY: misc_feature
248 <222> LOCATION: (1)..(32)
249 <223> OTHER INFORMATION: PCR Primer
252 <400> SEQUENCE: 7
253 gagaggatcc ttaggtgtca attttatcct gc 32
256 <210> SEQ ID NO: 8
257 <211> LENGTH: 29
258 <212> TYPE: DNA
C--> 259 <213> ORGANISM: Artificial
261 <220> FEATURE:
262 <221> NAME/KEY: misc_feature
263 <222> LOCATION: (1)..(29)
264 <223> OTHER INFORMATION: PCR Primer
267 <400> SEQUENCE: 8
268 gagacatatg ttatcagagc aggttcagc 29
271 <210> SEQ ID NO: 9
272 <211> LENGTH: 32
273 <212> TYPE: DNA
C--> 274 <213> ORGANISM: Artificial
276 <220> FEATURE:
277 <221> NAME/KEY: misc_feature
278 <222> LOCATION: (1)..(32)
279 <223> OTHER INFORMATION: PCR Primer
282 <400> SEQUENCE: 9
283 gagaggatcc ttaggtgtca attttatcct gc 32
286 <210> SEQ ID NO: 10
287 <211> LENGTH: 22
288 <212> TYPE: DNA
C--> 289 <213> ORGANISM: Artificial
```

See P. 3

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/830,026

DATE: 05/07/2001
 TIME: 14:25:59

Input Set : A:\Uok532-1.txt
 Output Set: N:\CRF3\05072001\I830026.raw

```

291 <220> FEATURE:
292 <221> NAME/KEY: misc_feature
293 <222> LOCATION: (1)..(22)
294 <223> OTHER INFORMATION: PCR Primer
297 <400> SEQUENCE: 10
298 gagacatatg ttgcaaaaagc aa                22
301 <210> SEQ ID NO: 11
302 <211> LENGTH: 29
303 <212> TYPE: DNA
C--> 304 <213> ORGANISM: Artificial
306 <220> FEATURE:
307 <221> NAME/KEY: misc_feature
308 <222> LOCATION: (1)..(29)
309 <223> OTHER INFORMATION: PCR Primer
312 <400> SEQUENCE: 11
313 gagactcgag atgcgttttt ttggcaccg          29
316 <210> SEQ ID NO: 12
317 <211> LENGTH: 29
318 <212> TYPE: DNA
C--> 319 <213> ORGANISM: Artificial
321 <220> FEATURE:
322 <221> NAME/KEY: misc_feature
323 <222> LOCATION: (1)..(29)
324 <223> OTHER INFORMATION: PCR Primer
327 <400> SEQUENCE: 12
328 gagactcgag acccagagaa gaacttacg          29
331 <210> SEQ ID NO: 13
332 <211> LENGTH: 30
333 <212> TYPE: DNA
C--> 334 <213> ORGANISM: Artificial
336 <220> FEATURE:
337 <221> NAME/KEY: misc_feature
338 <222> LOCATION: (1)..(30)
339 <223> OTHER INFORMATION: PCR Primer
342 <400> SEQUENCE: 13
343 gagaggatcc ttaagctcga atgttaccag          30
346 <210> SEQ ID NO: 14
347 <211> LENGTH: 27
348 <212> TYPE: DNA
C--> 349 <213> ORGANISM: Artificial
351 <220> FEATURE:
352 <221> NAME/KEY: misc_feature
353 <222> LOCATION: (1)..(27)
354 <223> OTHER INFORMATION: PCR Primer
357 <400> SEQUENCE: 14
358 gagacatatg ttgcaaaaagc aatttgc          27
361 <210> SEQ ID NO: 15
362 <211> LENGTH: 31
363 <212> TYPE: DNA

```

VERIFICATION SUMMARY
PATENT APPLICATION: US/09/830,026

DATE: 05/07/2001
TIME: 14:26:00

Input Set : A:\Uok532-1.txt
Output Set: N:\CRF3\05072001\I830026.raw

L:10 M:270 C: Current Application Number differs, Replaced Current Application No
L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:184 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:3
L:199 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:4
L:214 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:5
L:229 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:6
L:244 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:7
L:259 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:8
L:274 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:9
L:289 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:10
L:304 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:11
L:319 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:12
L:334 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:13
L:349 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:14
L:364 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:15
L:379 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:16
L:394 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:17